

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Jan. 16, 2003, 16:58:02 ; Search time 601.571 Seconds  
(without alignments)  
122 402 Million coll updates/sec

Title:	95-09-856 070 1A
Perfect score:	25
Sequence:	1 EREK 5

Scoring table:	BLASmR2	
Xqapop	10.0	Xqapext 0.5
Yqapop	10.0	Yqapext 0.5
Fqapop	6.0	Fqapext 7.0
Delop	6.0	Delcext 7.0

Searched: 15154066 seqs, 8047744376 residues

Total number of bits satisfying oxygen parameters: 32309132

Minimum	DB seq	length: 0
Maximum	DB seq	length: 2000000000

Post-processing: Minimum Match 6%  
Maximum Match 100%  
Listing first 45 summaries

```

Command line parameters:
-MODE frame_pcn model -DEV xif
-cmd 1751pt1.spool/inst666070/rmat_14012004_155834_1r54/app_query last_1_1592
-DH-EXT -OFMT=fastap -SUFFIX=first -MINMAX=0.1 -LOCPCL=0 -LOCPEXT=0
-UNITS=bits -START=1 -END=1 -MAIKX=BITSUM62 -TRANS=human0_cdi -LIST=45
-DOCALIGN=200 -THR_SCORE_pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTEXT=pt0 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US065856070 -GNN 1 1 -522 -freq 14012004_155834_1r54 -NFILE=4
-NO_XLIFY -NO_MMAP -LAP=EUDEY -NEG_SCF_PBS=0 -WAIT -LNCN=134 -DEFTIMEOUT=120
-WARN_TIMEOUT=10 -THRRADS=1 -XGAP=10 -XGAPEXT=5 -XGAPEXT=7
-XGAP=10 -XGAPEXT=5 -DELTOP=6 -DELEXT=7

```

```

Database :
EST : *
1 : em_estba : *
2 : em_esthum : *
3 : em_estim : *
4 : em_estimv : *
5 : em_estlmv : *
6 : em_estlpl : *
7 : em_estlro : *
8 : em_hlc : *
9 : gb_est1 : *
10 : gb_est2 : *
11 : gb_hlc : *
12 : gb_est3 : *
13 : gb_est4 : *
14 : gb_est5 : *
15 : em_estfun : *
16 : em_estfom : *
17 : gb_gss : *
18 : em_gss_hum : *
19 : em_gss_inv : *
20 : em_gss_fin : *
21 : em_gss_vrt : *
22 : em_gss_fur : *
23 : em_gss_mam : *
24 : em_gss_mrg : *
25 : em_gss_oth : *
26 : em_gss_pro : *
27 : em_gss_roo : *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query	Score	Query			L1	Description
			Match	Length	LB		
C 1	25	100.0	31	17	A2438877	A2438877 IM0212897	
C 2	25	100.0	37	17	A2336778	IM0067611	
C 3	25	100.0	43	17	A2487023	IM0069404	
C 4	25	100.0	40	17	A2511782	IM0356A24	
C 5	25	100.0	42	17	A2817485	IM0086118	
C 6	25	100.0	48	17	AL754610	ArabIdops	
C 7	25	100.0	49	17	A2827382	IM0103113	
C 8	25	100.0	50	9	AU107971	AU107971	
C 9	25	100.0	50	9	AU107973	AU107973	
C 10	25	100.0	51	17	A2548690	IM0498A17	
C 11	25	100.0	52	10	AW632145	HF052F09S	
C 12	25	100.0	52	17	HR862110	SAFV_9887	
C 13	25	100.0	53	17	A2345637	IM0048A11	
C 14	25	100.0	57	17	A2996539	IM0282F14	
C 15	25	100.0	55	9	AL683110	FW33306_X	
C 16	25	100.0	55	17	A2327114	IM0850J13	
C 17	25	100.0	55	17	A2494560	IM0267N14	
C 18	25	100.0	57	17	A2811159	IM0307K10	
C 19	25	100.0	57	13	RI736466	FW360112	
C 20	25	100.0	54	14	FW479185	KU28e02_Y	
C 21	25	100.0	58	17	A2766334	IM0563017	
C 22	25	100.0	59	10	AW694582	FW077H04S	
C 23	25	100.0	60	17	HH791807	SAIK_0614	
C 24	25	100.0	60	17	HH791808	SALK_0614	
C 25	25	100.0	64	17	A2363047	IM0108322	
C 26	25	100.0	65	10	EE579001	FW06K04_Y	
C 27	25	100.0	65	17	A2609314	IM0434F02	
C 28	25	100.0	65	13	HI430347	IM0434F02	
C 29	25	100.0	67	17	HH811544	IM0434F02	
C 30	25	100.0	68	9	AA792541	IM0434F02	
C 31	25	100.0	68	17	AW695648	IM0967316	
C 32	25	100.0	70	17	HH414753	IM0739F0	
C 33	25	100.0	70	10	HE059656	FW35D07_Y	
C 34	25	100.0	71	17	A2779808	IM0016A12	
C 35	25	100.0	72	9	AL311947	IM0016A12	
C 36	25	100.0	72	9	AL802430	IM0016A12	
C 37	25	100.0	73	10	AW636548	SWV_L3CAN	
C 38	25	100.0	73	17	AL770284	ArabIdops	
C 39	25	100.0	74	10	AW641269	IM096D02_W	
C 40	25	100.0	75	10	FW324195	FW016D04P	
C 41	25	100.0	75	12	FW575055	UTSWL_S08H	
C 42	25	100.0	77	17	A2825790	IM0212897	
C 43	25	100.0	79	17	A2661575	IM0540B11	
C 44	25	100.0	82	13	BI420680	LJNEST60C	
C 45	25	100.0	82	13	HM300131	IM0540B11	

## ALIGNMENTS

RESULT 1	AZ428877	21 bp	1NA	linear	Q53-03-0CT-2000			
LOCUS	AZ428877/c	1046	plasmid	100GCM	library Mus musculus genomic			
DEFINITION	clone U03C10U212N07 R. DNA sequence.							
ACCESSION	AZ428877							
VERSION	AZ428877.1	GI:10552986						
KEYWORDS	GSS.							
SOURCE	house mouse.							
ORGANISM	Mus musculus							
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.							
REFERENCE	1 (bases 1 to 21)							
AUTHORS	Punn, P., Aoyagi, A., Barber, M., Heacorn, F., Buval, B., Hamil, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly							

M., Rose M., Pose P., Stokes P., Tinney A., von Niederhausern A., and Wright D., Weiss R.  
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 304, Biomedical Polymers Research Bldg., 20 S. 2040 E., SIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0212 row: N column: 07  
 Seq primer: CACACAGCAAAACACCATGACC  
 Class: plasmid ends  
 High quality sequence stop: 21.  
 Location/Qualifiers  
 1..21

## FEATURES

source

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="M0067N11"  
 /clone\_lib="Mouse 10kb plasmid M0067N11 library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, TI-resistant, P-"  
 /note="Vector: pMD20nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adapter oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g1473214/gb1AF124072.1), a copy number inducible derivative of plasmid p1, the vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
 0 a 6 c 0 g 15 t

Alignment Scores:  
 Pred. No.: 2,940-03 Length: 21  
 Score: 25.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 17 Gaps: 0

US-09-856-070-16 (1-5) x AZ428877 (1-21)

QY 1 GluArgGluLysGlu 5  
 Db 15 GAAAGAGAGAGAGAA 1

## RESULT 2

AZ436778/c

LOCUS AZ436778 47 bp DNA Linear GSS 29-SEP-2000  
 DEFINITION M0067N11F Mouse 10kb plasmid M0067N11 F. DNA sequence.  
 clone M0067N11 F. DNA sequence.

ACCESSION AZ436778

VERSION AZ436778.1 G1:10406421

KEYWORDS GSS.

SOURCE house mouse

ORGANISM Mus musculus

## REFERENCE

AUTHORS

## TITLE

JOURNAL

COMMENT

Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;  
 Mammalia, Eutheria, Rodentia, Sciurognathi; Muridae; Mus; Mus.  
 1 (bases 1 to 37)  
 Dunn,D., Aoyagi,A., Barber,M., Heacorn,L., Duval,H., Hamill,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,L., Reilly,  
 M., Rose,M., Pose,P., Stokes,P., Tinney,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 304, Biomedical Polymers Research Bldg., 20 S. 2040 E., SIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0067 row: N column: 11  
 Seq primer: CGTGTGAAAACGACGCCCAT  
 Class: plasmid ends  
 High quality sequence stop: 37.  
 Location/Qualifiers  
 1..37

## FEATURES

source

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="M0067N11"  
 /clone\_lib="Mouse 10kb plasmid M0067N11 library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, TI-resistant, P-"  
 /note="Vector: pMD20nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adapter oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g1473214/gb1AF124072.1), a copy number inducible derivative of plasmid p1, the vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
 8 a 12 c 1 g 16 t

Alignment Scores:  
 Pred. No.: 4,590-03 Length: 47  
 Score: 25.00 Matches: 5  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 17 Gaps: 0

US-09-856-070-16 (1-5) x AZ336778 (1-37)

QY 1 GluArgGluLysGlu 5

Db 23 GAAAGAGAGAGAGAA 9

## RESULT 3

AZ987023/c

LOCUS

DEFINITION

AZ987023 39 bp DNA Linear GSS 27-APR 2001  
 M0026924F Mouse 10kb plasmid M0026924 F. DNA sequence.  
 clone M0026924 F. DNA sequence.

```

ACCESSION      AZ987023
VERSION        A7987023.1  GI:13958250
KEYWORDS       GSS
SOURCE         house mouse,
ORGANISM       Mus musculus
REFERENCE      Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
AUTHORS       Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus,
              Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
              Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Peilly
              M., Pose, M., Pose, P., Stokes, P., Tingey, A., von Niederhausern, A.,
              and Wright, D., Weiss, R.
TITLE         Mouse whole genome scaffolding with paired end reads from 10kb
              plasmid inserts
JOURNAL        Unpublished (2000)
COMMENT       Contact: Robert R. Weiss
              University of Utah Genome Center
              University of Utah
              Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
              84112, USA
              Tel: 801 585 5606
              Fax: 801 585 7177
              Email: ddunn@genetics.utah.edu
              Insert length: 10000 Std Error: 0.00
              Plate: 0269 row: N column: 24
              Seq primer: CGTTCTAAACGACGCCAGT
              Class: plasmid ends
              High quality sequence stop: 39
              Location/Qualifiers
FEATURES       1..39
              source          /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="mimc2M036Q24"
              /sex="Female"
              /lab_host="E. coli strain XL10-Gold, 11-resistant, P-"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adaptor DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of pWD42 (g1147321141gblAF129072.1), a copy-number
              inducible derivative of plasmid R1. The vector was ligated
              with adaptors complementary to the insert adaptors and
              purified. The sheared, adaptor mouse DNA was annealed to
              adaptor vector DNA, and transformed into
              chemically competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
BASE COUNT     0 a 17 c 1 g 21 t
ORIGIN
Alignment Scores:
Pred. No.:      4,9e+03      Length:      39
Score:          25,00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             17      Gaps:      0

US-09-856-070-16 (1-5) x AZ987023 (1-39)
QY 1 GluArgGluTysGln 5
   |||||
DB 23 GACAGAGAGAGAG 9

```

```

RESULT 4
ACCESSION      AZ511352
VERSION        A511352.1  GI:13692668
KEYWORDS       GSS
SOURCE         house mouse,
ORGANISM       Mus musculus
REFERENCE      Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
AUTHORS       Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus,
              Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
              Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Peilly
              M., Pose, M., Pose, P., Stokes, P., Tingey, A., von Niederhausern, A.,
              and Wright, D., Weiss, R.
TITLE         Mouse whole genome scaffolding with paired end reads from 10kb
              plasmid inserts
JOURNAL        Unpublished (2000)
COMMENT       Contact: Robert R. Weiss
              University of Utah Genome Center
              University of Utah
              Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
              84112, USA
              Tel: 801 585 5606
              Fax: 801 585 7177
              Email: ddunn@genetics.utah.edu
              Insert length: 10000 Std Error: 0.00
              Plate: 0356 row: A column: 24
              Seq primer: CGTTCTAAACGACGCCAGT
              Class: plasmid ends
              High quality sequence stop: 40
              Location/Qualifiers
FEATURES       1..40
              source          /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="mimc1M0356A24"
              /sex="Male"
              /lab_host="E. coli strain XL10-Gold, 11-resistant, P-"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adaptor DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of pWD42 (g1147321141gblAF129072.1), a copy-number
              inducible derivative of plasmid R1. The vector was ligated
              with adaptors complementary to the insert adaptors and
              purified. The sheared, adaptor mouse DNA was annealed to
              adaptor vector DNA, and transformed into
              chemically competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
BASE COUNT     0 a 20 c 0 g 20 t
ORIGIN
Alignment Scores:
Pred. No.:      5e+03      Length:      40
Score:          25,00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             17      Gaps:      0

US-09-856-070-16 (1-5) x AZ511352 (1-40)

```

```

QY      1 GluArqGluLysGlu 5
DB      21 GAGACAGAGAGAGAG 7

RESULT 5
A2817465
LOCUS   2M00861196 Mouse 10kb plasmid YUGCLM library Mus musculus genomic
DEFINITION
ACCESSION A2817465
VERSION A2817465.1 GI:12987289
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, R., Hamil, C.,
Islam, H., Lougare, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, F., Stokes, P., Tinney, A., von Nischta-Hausen, A.
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0086 row: 1 column: 18
Seq primer: CACACAGGAAACACATGACC
Class: plasmid ends
High quality sequence stop: 42.
FEATURES
source
location/Qualifiers
1..42
/organism="Mus musculus"
/strain="C57BL/6T"
/db_xref="taxon:10090"
/clone="UUGC2M0086118"
/clone_lib="Mouse 10kb plasmid YUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pMD20v; Purified genomic DNA from Mus
musculus C57BL/6T (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD2 (ql1472124) (pAF129072.1), a copy-number
inducible derivative of plasmid p1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 21 a 1 c 15 g 5 t
ORIGIN

Alignment Scores:
Pred. No.: 5,82e-03 Length: 42
Score: 25.00 Matches: 5
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 17 Gaps: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-09-856-070-16 (1-5) x A2817465 (1-42)

QY      1 GluArqGluLysGlu 5
DB      23 GAGAGAGAGAGAGAG 37

RESULT 6
AL754610
LOCUS   Arabidopsis thaliana T-DNA flanking sequence GK-054H11-012301,
DEFINITION genomic survey sequence.
ACCESSION AL754610
VERSION AL754610.1 GI:21487108
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE
AUTHORS Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Siedler, H.
and Weisshaar, B.
TITLE A pipeline for automated high-throughput generation of FSLs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
JOURNAL Unpublished
REFERENCE
AUTHORS Rosso, M., Strizhov, N., Li, Y., Peiss, B., Dekker, K. and Weisshaar, B.
TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
JOURNAL Unpublished
REFERENCE
AUTHORS Rosso, M., Strizhov, N., Li, Y. and Weisshaar, B.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weisshaar B., Max-Planck Institut fuer
Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion close to or within gene Atg51520. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES
source
location/Qualifiers
1..48
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-054H11-012401"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/notes="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector pAC161. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"
BASE COUNT 0 a 19 c 3 g 26 t
ORIGIN

Alignment Scores:
Pred. No.: 5,82e-03 Length: 48
Score: 25.00 Matches: 5
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 17 Gaps: 0

```



```

SOURCE
ORGANISM      human.
Eukaryota: Mollusca; Chordata, Vertebrata, Euteleostomi;
Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS      1 (bases 1 to 50)
              Suzuki, Y., Iwata, H., Tsukada, T., Mizushima Sugano, J., Suter, J., Hata
              H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., Okubo, K., Sakaki
              Y., Nakamura, Y., Suyama, A., and Sugano, S.
TITLE        Diverse transcriptional initiation revealed by fine, large-scale
              mapping of mRNA start sites
JOURNAL      EMBO Rep. 2 (5), 386-393 (2001)
MEDLINE      21276072
COMMENT      Contact: Yutaka Suzuki
              Department of Virology
              Institute of Medical Science, University of Tokyo
              4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
              Email: shizuka@ims.u-tokyo.ac.jp
              Suzuki, Y., Yoshimoto-Nakagawa, K., Maruyama, K., Suyama, A., and Sugano
              S. Construction and characterization of a full length enriched and
              a 5'-end-enriched cDNA library Gene 260 (1-2): 149-156 (1997)
FEATURES
source       1..50
              Location/Qualifiers
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="ZRR6C748"
              /note="Differential display comparison of untreated and
              dimethylsulphate treated U937 cells"
BASE COUNT   4 a 19 c 7 q 20 t
ORIGIN
Alignment Scores:
  Pred. No.:      6,02e+03      Length:      50
  Score:          25.00         Matches:      5
  Percent Similarity: 100.00%   Conservative: 0
  Best Local Similarity: 100.00% Mismatches:    0
  Query Match:     100.00%     Indels:       0
  DB:              9           Gaps:         0

US-09-856-070-16 (1-5) x A0137973 (1-53)

QY      1 GluAraGluLysGlu 5
        |||||
DB      17 GAGACGCAAAACAG 3

RESULT 10
AZ638690
LOCUS      AZ638690.1 51 bp DNA linear GSS 13-DEC-2000
DEFINITION clone U0GCLM0498A17 R, DNA sequence.
ACCESSION  AZ638690
VERSION     1
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.
Eukaryota: Mollusca; Chordata, Vertebrata, Euteleostomi;
Mammalia: Eutheria; Rodentia; Scuriognathi; Muridae; Mus
1 (bases 1 to 51)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmood, M., Meenen, K., Pedersen, T., Reilly
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert H. Weiss
University of Utah Genome Center
University of Utah
Em. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

human.
Insert Length: 10000 Std Error: 0.00
Plate: 0498 row: A column: 17
Seq primer: CAACAGCAAAACAGCAGACC
Class: plasmid ends
High quality sequence stop: 51.
Location/Qualifiers
1..51
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U0GCLM0498A17"
/clone_lib="Mouse 10kb plasmid U0GCLM library"
/sex="Male"
/lab_host="F. Coli strain XL10-Gold, TI-resistant, F-"
/notes="Vector: pMD42uv, Fufilled genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/Instruments/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.205 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi:14732114|gb|AF129072.1), a copy number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT   1 a 28 c 0 q 22 t
ORIGIN
Alignment Scores:
  Pred. No.:      6,12e+03      Length:      51
  Score:          25.00         Matches:      5
  Percent Similarity: 100.00%   Conservative: 0
  Best Local Similarity: 100.00% Mismatches:    0
  Query Match:     100.00%     Indels:       0
  DB:              17           Gaps:         0

US-09-856-070-16 (1-5) x AZ638690 (1-51)

QY      1 GluAraGluLysGlu 5
        |||||
DB      15 GAAACAGCAAAACAG 1

RESULT 11
AW692145/c
LOCUS      AW692145/c
DEFINITION NF052F09STF1000 Developing stem Medicago truncatula cDNA clone
ACCESSION  AW692145
VERSION     1
KEYWORDS    EST.
SOURCE      barrel medic.
ORGANISM    Medicago truncatula
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta, Magnoliophyta, eudicotyledons; core eudicots;
Fosidae; eurosids I; Fabales; Fabaceae; Papilionoideae, Trifoliaceae;
Medicago.
1 (bases 1 to 52)
Re, X.-Z., Shadie, G., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell
C.J., Flores, H.R., Imman, J.T., Weller, J.W., May, G.D., and Dixon
R.A.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula stem library
Unpublished (2000)
On Apr 14, 2003 this sequence version replaced gi:7566881.
Contact: Dixon RA

```

Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 421 7202  
Fax: 580 221 7380  
Email: radixon@noble.org  
Insert Length: 726 Std Error: 0.00  
Plate: 052 row: F column: 09  
Seq primer: TCACACAGGAGAGGATATGAC.

## FEATURES

source

1..52  
Location/Qualifiers  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF052P0951"  
/clone\_lib="Developing stem"  
/tissue\_type="stem"  
/dev\_stage="Pooled developmental"  
/note="Vector: Lambda Zap; Contains a mixture of  
intermodal stem segments"

BASE COUNT 8 a 20 c 2 g 22 t

## ORIGIN

Alignment Scores:  
Pred. No.: 6,22e-03 Length: 52  
Score: 25.00 Matches: 5  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-856-070-16 (1-5) x AW692145 (1-52)

QY 1 GluArgGluLysGlu 5

Db 33 GAAGAGGAGAGAGAG 19

RESULT 12

HH862110/C

LOCUS

DEFINITION

Arabidopsis thaliana DNA insertion lines

Arabidopsis

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA. This sequence lies within an annotated exon of At1g64866.

Class: TDNA tagged.

location/Qualifiers

1..52

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="SALK\_088796"

/clone\_lib="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more DNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

BASE COUNT 11 a 12 c 8 g 21 t

## ORIGIN

Alignment Scores:

Pred. No.: 6,22e-03 Length: 52  
Score: 25.00 Matches: 5  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 17 Gaps: 0

US-09-856-070-16 (1-5) x HH862110 (1-52)

QY 1 GluArgGluLysGlu 5

Db 29 GAACGGAGAGAGAG 15

RESULT 13

AZ325637/C

LOCUS

DEFINITION

Mouse 10kb plasmid UMGCLM library

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

Unpublished (2000)

Contact: Robert H. Weiss

University of Utah

Genome Center

84112, USA

Tel: 801 585 5800

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0048 row: A column: 11

Seq primer: CGTGTGAAAAGAGGCGGACG

Class: plasmid ends

High quality sequence stop: 53.

location/Qualifiers

1..53

/organism="Mus musculus"

/strain="C57Bl/6J"

/db\_xref="taxon:10090"

/clone="UMGCLM0048A11"

/clone\_lib="Mouse 10kb plasmid UMGCLM library"

/sex="Male"

/lab\_host="E. coli strain XL10-Gold, Ti-resistant, F-"

/note="Vector: FWD42nv; Purified genomic DNA from M.

musculus C57Bl/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/document/s/hares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and 14

polynucleotide kinase. Adapter oligonucleotides were

ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (G14732114[6]AP129972.1); a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

```

BASE COUNT      a  24 c  1 g  28 t
ORIGIN
Alignment Scores:
Pred. No.:      6,41c-03      Length:      53
Score:          25.00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              17      Gaps:      0

US 09 856-070-16 (1-5) x AZ996537 (1-53)

QY 1 GluAraqluLyslu 5
DB 22 CACACAGACAAAGAA 8

RESULT 14
LOCUS AZ996539/c 54 bp DNA linear GSS 27-APR-2001
DEFINITION 2M02B2F14R Mouse 10kb plasmid 39532M library Mus musculus genomic
clone U00C2M02B2F14 R. DNA sequence.
ACCESSION AZ996539
VERSION AZ996539.1 GI:13867766
KEYWORDS mouse mouse;
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 54)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Lonnacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 408, Biomedical Polymers Research Bldg., 20 S. 2030 E., St.C, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 6282 row: F column: 14
Seq primer: CACACAGACAAAGCAATGACC
Class: plasmid ends
High quality sequence step: 54.
FEATURES
Location/Qualifiers
1..54
/organism="Mus musculus"
/strain="G57H1/bj"
/db_xref="taxon:10690"
/clone="U00C2M02B2F14"
/clone_lib="Mouse 10kb plasmid U00C2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, II-resistant, F-"
/vector="Vector: pMD42v; Purified genomic DNA from M.
musculus G57BL/6J (female) was obtained from the Jackson
Laboratory Mouse DNA Resource

```

(http://www.taxidb.org/taxidb.cgi). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt ended with 14 DNA polymerase and 14 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (G14732114[6]AP129972.1); a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

```

BASE COUNT      a  21 c  5 g  27 t
ORIGIN
Alignment Scores:
Pred. No.:      6,41c-03      Length:      54
Score:          25.00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              17      Gaps:      0

US-09-856-070-16 (1-5) x AZ996539 (1-54)

QY 1 GluAraqluLyslu 5
DB 34 CACACAGACAAAGAG 20

RESULT 15
LOCUS A1689110 55 bp mRNA linear EST 16-DEC-1999
DEFINITION LX92306.x1 NCI-CGAP_U14 Homo sapiens cDNA clone IMAGE:2277010 4'
similar to TR-064657 Q64657 INTEGRIN BETA 5 SUBUNIT 1, mRNA
sequence.
ACCESSION A1689110
VERSION A1689110.1 GI:4900404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 55)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.livni.gov/bhrp/image/image.html

```

Trace considered overall poor quality

Insert length: 1515 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence step: 1.

Location/Qualifiers

1..55

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2277010"

/clone\_lib="NCI-CGAP\_U14"

/tissue\_type="serous papillary carcinoma, high grade, 2



pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; Salt:  
Site 2: Not cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.48 kb. Life Technologies catalog #:  
11542-016"

HASE COUNT 31 a 11 c 13 g 0 t

ORIGIN

Alignment Scores:

Pred. No.:	6.51e+03	Length:	55
Score:	25.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	0
Best local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-856-070-16 (1-5) x A1689110 (1-55)

QY 1 GluArgGluIysGlu 5

Db 33 GAAAGAGAGAAAGAA 47

Search completed: January 16, 2003, 21:36:58  
Job time : 664.571 secs

